



Numèro de publication:

0 360 677 Δ1

DEMANDE DE BREVET EUROPEEN **a** 

Numéro de dépôt: 89402549.3

® Int. CI.5 C 12 Q 1/68

@ Date de depôt 18.09.89

31/33, rue de la Fédération

@ Priorité. 20.09.88 FR 8812285

 Date de publication de la demande 28.03.90 Bulletin 90/13

Etats contractants désignés CH DE ES FR GB LI NL SE

Demandeur: COMMISSARIAT A L'ENERGIE ATOMIQUE F-75015 Paris (FR)

(72) Inventeur: De Cosnac, Bertrand 11 Bis, avenue Mac Mahon F.75017 Paris (FR)

Grimm, Bartrand F-92295 Chatenay Malabry Cedex (FR)

Haan, Serge 18 Domaine de la Butte é la Reine F-91120 Pelaiseau (FR)

(14) Mandataire, Lhoste, Cetherine et al BREVATOME 25, rue de Ponthieu F-75008 Paris (FR)

Procédé et installation d'identification des bases d'un ADN.

 Le procédé de l'invention comprend essentiellement une etape (2) de fragmentation de l'ADN, une étape (4) de reactions chimiques spécifiques des bases pour former des familles de cheines de taille différente, chaque famille comportant à une entremes de tante umerente, chaque tamine comportant a une extrémité une base spécifique, une étape (6) de marquage spécifique desoites familles de chaînes et de marquage de appenique depontes names de chaîne avec des isotopes ou de cheque nucléotide de chaque chaîne avec des isotopes ou de petites molécules stables, une étape d'introduction une à une des chaînes marquees dans une source d'ions, une êtepe (10) d'etomisation et d'ionisation de chaque chaîne marquée avec la source d'ions, une étape (12) de détection en temps réel des ions formés par spectromètre de masse, et une étape (14) de treitement des signaux issus de la spectrometrie de masse, ces signaux ètent représentetifs de la longueur des chaînes et de leur marquege spécifique

Upper(1) IX (44) IV(1)	-	professor that IE MANA
CHIEF MARKET	1945/17	Married Arrest
	}. }.	
HIMTON WA	] z ] a	PIG. 1

Bundesorvewere: Berin

### Description

# PROCEDE ET INSTALLATION D'IDENTIFICATION DES BASES D'UN ADN.

L'invention se rapporte à un procèdé et une installation d'identification des motifs d'une molècule organique complexe et en particuler d'une molècule biologique complexe telle qu'un acide nucleique , une proteine, un peptide, un polymère. Elle s'applique dans les domaines du gênie génétique, de la biologie, de la pharmacologie, tant au stade de l'expérimentation, de la recherche fondamentale qu'industrielle

Le procede de l'invention s'applique de façon generale à toute molecule constituée d'une succession d'acides amines, de nucléotides, de groupes d'atomes et, plus spécialement, aux acides nuclètiques, tels que

l'ADN (acide désoxyribonucléique) et l'ARN (acide ribonucleique). L'identification des sequences ou de l'ordre des nucléofides de l'ADN ou l'ARN est connue sous le nom de

Le séquençage des acides nucléiques a pris une réelle importance dans le domaine de la biologie moleculaire et de la biotechnologie. En particulier, le séquencage de l'ADN permet l'analyse des génomes animaux, végétaux et virsux dans des génes particuliers. Puisque la fonction d'une molécule biologique est déterminée par sa structure, la définition de la structure d'un gêne est cruciale pour la manipulation éventuelle

de cette unité de base d'information. On rappelle qu'un nucléotide est formé par une basé, un sucre et un phosphate.

Pour les acides nucleiques, les bases sont des bases azotées possèdant une structure héterocyclique du type purique ou pyrimidique , elles sont au nombre de 4. Le sucre est le ribose pour l'ARN et le desoxyribose pour l'ADN et le phosphate l'acide phosphorique.

Generalement, les nucleotides des acides nucleiques utilisés sont triphosphatés car c'est ce type de molècules qui fournit l'énergie necessaire au fonctionnement des enzymes dans de nombreuses réactions biochimiques, en particulier celles intervenant dans la polymerisation des ADN.

L'ARN et l'ADN présentent trois bases communes: l'adénine (A) et la guanine (G) qui sont des purines et la cytosine (C) qui est une pyrimidine. L'autre base pyrimidique est la thymine (T) pour l'ADN et l'uracile (U) pour

Pour l'ADN, les quatre types de nucléotides concernés sont le désoxy adénosine triphosphate ou dATP, le desoxy guarine triphosphate ou dGTP, le desoxy cytosine triphosphate ou dCTP et le desoxy thymine triphosphate ou dTTP. On les note ci-après dXTP avec X = A, G, C ou T

L'ADN contient l'information génétique complète nécessaire à la fabrication des protéines, éléments constitutifs de tout organisme vivant. En particulier, chaque succession de trois bases est le code par l'intermediare des ARN messagers, d'un acide aminé qui entre dans la composition d'une proteine. Le sequençage a pour but d'obtenir l'ordre des bases constituant un brin d'ADN ou d'ARN

Les methodes de sequençage de l'ADN comportent essentiellement 5 étapes qui sont (a) la fragmentation du genome. (b) les reactions chimiques de sequençage pour former des chaînes de nucléotides. (c) le marquage des chaines de nucleotides, (d) la separation des chaînes marquees et (e) la detection

45

ريس مرادي . Un genome ou ensemble des secteurs de l'ADN est constitué par un grand nombre de paires de bases (13 x 10° pour la levure et 3 x 10° pour l'homme). Le génome du fait de sa longueur, ne peut être directement. soumis aux reactions de sequençage. L'ADN est scindé en fragments d'environ 1000 à 30000 bases qui sont ensuite multiplies par cionage. Par exemple, cette tragmentation peut être effectuée par l'utilisation d'enzymes de restriction qui coupent l'ADN au niveau de certaines sequences ou par l'utilisation des uttra-sons comme expose dans l'article "Current Approaches to DNA Sequencing" de J.C. Moores Analytic Biochemistry 163, 1-8 (1987)

Elle consiste à appliquer aux chaînes obtenues en (a) quatre sèries de réactions chimiques spécifiques des quatre bases de l'ADN. Ces réactions conduisent à l'obtention de quatre familles de chaînes d'ADN dont l'une des extremités est communement fixée et l'autre est constituée d'une base spécifique de l'une des quatre

Deux methodes sont communément utilisées pour obtenir les 4 familles de chaînes d'ADN ci-dessus. L'une. senes de réaction developée par Maxam et Gilbert (Maxam A.M. et Gilbert W., Proc. Natl. Acad. Sci., U.S.A., vol. 74, № 2, pp 550-554 de 1977, "A new method for sequencing DNA") fait appel à une procédure de dégradation chimique des chaînes et l'autre, introduite par Sanger (Sanger F, et al., Proc. Natl. Acad. Sci. USA 74, pp. 5453-5467 de Decembre 1977 "DNA sequencing with chain-terminating inhibitors") est une méthode enzymatique et

emploie un procédé de terminaison de chaînes didésoxy. Dans la methode de Sanger, chacune des quatre réactions de séquençage met en jeu un grand nombre d'exemplaires d'ADN monocaténaire à sequencer , un amorceur (ou en terminologie anglo-saxonne primer) qui est un acide-nucléique complementaire de l'une des extrémités de l'ADN à sequencer ; les qualte nucleorides dXTP ; une enzyme qui permet la fabrication du brin complémentaire de l'ADN a sequencer, à partir des nucleotides et en commençant par l'amorceur et enfin un seul des quatre di-désoxy nucleotides appele nucléotide ferminal et note ddXTP. Ce ddXTP est un nucléotide auquel il manque le groupement OH du

arbone 3' du désoxyribose. L'amorceur se lie à l'Abn. à séquencer par complémentarité puis l'enzyme synthètise le brin L'amorceur se se à l'AUN à sequencer par complementante puis l'enzyme synthesise le brin. complementante jusqu'à ce qu'elle incorpore le ddXTP arrêtant (élongation de la chaîne. L'incorporation du carbone 3' du désoxyribose complementaire jusqui a ce qui ese inicorpore se color in altreusis sebrigaturo de la chaire. Unicorporazion du nucléotide terminali peut se produre n'importe quand au cours de la synthèse ce qui se traduit, à la fin de nuceotitie reminiai peur se procure n'importe quaro au cours de la synthése ce qui se traduit, a la nn de l'opération de sequençage spécifique d'une base X donnée, car la présence d'une familie d'ADN représentant

, пречания им sequençage specinique d'une base à donnée, par la presence d'une tam toutes les longueurs possibles dépuis l'amorceur jusqu'à environ 400 nucleotides ures les longueurs possibles depuis l'amorceur jusqu'à enveron aux nucleoides. Appliquée en parallèle pour les quatre ddXTP, la sêrie de quatre réactions conduit à un ensemble de quatre Appriques en parames pour les quairs qua.in-, la sene de quairs (ractions consuit à un ensentide de quairs familles de chaînes d'ADN ; les chaînes de chaque famille étant représentatives d'une base X gonnée sont

<sub>Naurie</sub> deposées sur un gei d'electrophorese. Dans la méthode de Maxam et Gibert, des chaînes d'ADN de dimensions adaptées et analogues à celles Lians is metrode de Maxam et usidert, des chaines d'ALPI de dimensions adspitées et analogues à celles Issues du procédé de Sanger sont produites par un tratement chimique d'ADN purifie. Cette operation permet issues ou procede de sanger sont produites par un traitement crimique d'auxi puriné. L'ette operation permét de séquencer un ADN mono- ou bicatenare marqué à vez un phosphate radioactif à l'une de ses extrémités.

e sequencer un AUN mono- ou picaienare marque avec un pnospinate radiciacist a l'une de ses extremités. Pour chacune des quatre séries de réaction de séquencage, l'ADN est soumis à un jeu de deux réactions Four chaculte des quarte series de resculuir de sequençage, i Aunt est soumes a un jeur de deux rescultis chimiques. Is première réaction chimique modifié et retire la base de son sucre, la seconde coupe l'ADN en crimiques, a première reaction crimique muoine et reure la base de son suvre, la seconde coupe i ALIN en retrant le sucre anni expose. On obtent ainsi des chaînes commençant à un engroit fixé et se termaint par retirant le sucre ainsi expose. Un oblent ainsi des chaînes commençant à un endroit rixe et se terminant par une base connue. Les conditions de réaction sont ajustées de taçon à produire des chaînes de toutes les une base connue. Les conditions de réaction sont ajustées de taçon à produire des chaînes de toutes les une usse commue. Les conomons de reaction sont ajustees de racon ( longueurs possibles comportant typiquement de 1 à 400 nucléotides.

iage c Les chaînes obtenues en (b) sont marquées préalablement, ou au cours des réactions, avec des marqueurs Les chaînes obtenues en (b) sont marquèes préalablement, ou au cours des réactions, avec des marqueurs radioactifs. Les marqueurs radioactifs Les marqueurs radioactifs utilisés sont le 22P ou le 35S, le tritum, le 14C qui sont des émetteurs de

yvonnements pr. Dans le cadre de la méthode de Sanger, il est possible de marquer l'amorceur, les dXTP ou les ddXTP rayonnements B-

ape d Les chaînes d'ADN marquèes sont alors separees suivant leur longueur par électrophorese, les plus petites Les chaines u Auxi i nai quees sons eous separeires survent eou longueur per meccopplicates, les pubs penter migrant le plus vite et les plus grandes le plus lentement. Le gel d'électrophorées qui est en général un gel de imgrant is plus vire et les plus grandes le plus sentiement. Le get di electrophor ese qui est en gamera un get de polyacrylamide permet de differencer les molécules de longueur N et N + 1 avec N variant de 1 à 400 environ pysicrysamide periner de universitéer les moves de songéeur n'en n'et la rec'h vesient de l'ei. L'électrophorese dure de 4 a 12 heures suivant la longueur du gel qui est ensuite seché.

20

35

45

55

upe s Le gel est ensure placé contre un film sensible au rayonnement béta. Les marqueurs radioactifs incorporés Le gel est ensuite place contre un tim sensible au rayonnement deta. Les marqueurs radioscitis nocipores dans les chaînes d'ADN sont utilisés pour développer une image auto-radiographique de chaque trace uars les unantes u num soni, umbres pour ueverupper une intege auto-reding system de uneque ruse électrophorétique. L'auto-radiographe est développée après 12 heures ou plus d'exposition au rayonnement. europrovenque. L'automanicagnatrie est uevenuppre après ; cientres ou pris u exposition au rayonitement. La position relative des bandes qui apparaissent sur l'image automadiographique, au sein des quatre Le publicier relieure une uerure qui supprincipent sur rannes surviridung aprincipe, au sen des cyatre colories correspondant aux quatre familles de réactions, permet la détermination de la sequence des bases

o inture un unique. Apres la lecture, l'examen et l'interprétation de l'auto-radiographie, la séquence est entrée dans la mémoire

un ordinateur pour une analyse et un assem plage uterieurs. Ces deux méthodes de séquercage d'AON précédemment décrités sont largement utilisées et comportent. de l'ADN de départ d'un ordinateur pour une analyse et un assem blage ultérieurs Les seux mismours se sequençage à Aux precedemment decrites sont argement utilisées et component, chacune pluseurs variantes. Blen que très performantes et efficaces, des méthodes demandent une mise en crecure puseurs variances, pien que tres penormantes et emisures, ces memores uminimient une mise de George fres complexe qui comporte de séneux inconvenients. C'est en réalite un tres gros travail, delinat qui oeuvre tres complexe qui comporte de serieux inconvenients. L'est en reafre un tres gros tievau, deuteut, qui doit être extrémement blen standardise pour que la source de connaissances et la comparaison des

igrations electrophoretiques scient nables. Par alleurs, la lecture des electrophoréses des chaînes radioactives se fait après auto-radiographie. La a ameurs, la secture des electrophicheses des charnes abdolectives se latit après autorisations parties autorisations par la susse, une parfaite reproductibilité est de rigueur fant dans le développement de ces autorisationgrammes que aussi, une parfaite reproductibilité est de rigueur fant dans le développement de ces autorisations que migrations électrophorétiques soient fiables.

ans teur recture. En effet, l'image sur film d'une bande de gel radioactif est plus large que la bande elle-même et la on enet.; I mage sur min usum panue ur yer ramanin est plus ratig que la usunde mercinente et a comparason des positions des bandes entre les quarra pistes différentes de l'électrophorése (qui payent comparason des posicions des usindes entre les quatre justes cinterentes de l'encuropriciese (qui jeuve-induré un comportement non uniforme quant aux mobilités des chaînes) peut limiter la résolution inquire un comportentent non uniforme quant, aux moderes des chaines) peut inmitet la residurion effectivement observés sur l'auto-radiogramme et donc rédure la longueur de la séquence qui peut être lue à effectivement observee sur l'auto-radiogramme et donc redure la longueur de la sequence qui peut parir des gels Typiquement, de 200 à 400 bases peuvent être lues à partir d'un simple jeu de tracès

i ini wes yers i yawalenieni. We www.a awa wasasayeuweni eue aless a pau in diun asingke gevian diacess. De plus, les prégularités suiverant d'un tracé à un autre peuvent rendre difficile la lecture automatique des ue plus, les irregularités survenairs u un trace à un autre provens renoire dinicie la recure automatique des auto-raciogrammes par des téchniques d'imagene existantes (par exemple densionnaires à baisyages) car auto-radiogrammes par ues sectiniques o imagene existantes (per exemple censioniteires a biasys).
Toeli humain peut actuellement miseux discriminer des irrégulantés que ne le font les ordinateurs.

per numain peut scureirent nieux discriminer ces dregueires que ne le lium les unimises une interes. La nécessité d'un son intensé quant à l'analyse "manuelle" des auto-radiogrammes demande donc du Par alleurs. l'unisation de marqueurs radio actifs présente de sérieux incoménients quant à la manipulation rar ameurs, i umbanon de marqueurs radio actits présente de senieux noonvenents quains la manipulazion (système de protection complexe des opérateurs et décontamination périodique des installations), au temps et est sujette à des erreurs

layaleme de protection complexe des operateurs et décontamination persondue d' stockage des produits durant l'analyse et au stockage des déchets en fin d'analyse. Ceffin, in est pas possible d'observer les bandes au sen du gét pendant l'électrophorèse afin de contrôler cnin, il n'est pas pusoine o disserver les bances au sen ou yer pendant i electrophorèse ain de comiser Toperation en Cours, mais l'aut nécessairement terminer l'éléctrophorèse à un moment détermine et attendre

le développement de l'auto-radiogramme avant de pouvoir lire la séquence developpement de l'autoraziogramine avant de pouvoir de la sespertive. Aussi, pluseurs procédes ont déla envisagé la détection en temps réel des chaînes d'ADN pendant leur migration au cours de l'électrophorèse

Les uns s'intéressent par exemple à la détection du rayonnement 8- émis par les chaînes marquées au <sup>32</sup>P, Les uns similer essemi par exemple a la derection du rayonnement p-entis par les chaines marquest les autres s'interessent à la détection de l'émission de fluorescence induite par un faisceau laser

s autres simiteressent e la cettécnion de remission de horiessence induite par un taisceau asser Dans le premier cas, la détection en ligne (ou temps réel) des bandes radioactives au cours de uans le premier cas, la detection en ligne (ou temps réel) des bandes radioactives au cours de l'électrophorése peut permettre une certaine simplification, par une certaine réduction du nombre des electropriores peut perioreire une cetame amipinicaron, per une cerame reduction ou nombre des opérations à effectuer, pour l'ensemble du procéde. Cependent, s'aut tenir compte de la courte periode de operations a energiter, plous resistentiale out procedes, seperations, sizus tenir compile de la courte periode de demivire du proceptore 32 et par conséquent de l'instabilité des réactifs de marquage radioactif. En outre, demi-vie du prosprore uz et par consequent de l'instatbille des reactis de marquage radocacif. En outre, comme dans les méthodes plus traditionnelles de Sanger ou Maxam et Gilbert, il demeure que des problemes

ont poses par l'eximination et la manipulation des produits radioacins. En outre, le temps de passage des chaînes devant le détecteur étant très réduit, cela implique une plus sont posès par l'élimination et la manipulation des produits radioactifs. an outre, le temps de passage des chaines devant le détecteur étant tres réduit, cela implique une plus faible efficacité de collection des rayonnements, ce qui peut entraîner des problèmes de sensibiléte et de

esoluvion. La détection de l'émission de fluorescence induite par un fasceau laser a été proposée par Ansorge (vor DE-A-3 618 605 et l'article dournit of Biochemical and Biophysical Methods. 13(1986), pp. 315-323. "A DEIAN 510 000 et l'article Journal of biochemical and biophysical methods. 13(1986), pp. 315-323. A nontradioactif automated method for DNA sequence determination\*). Cette methods utilise un seul

fluorophore pour marquer l'amorceur (ou primer) dans le cadre de la méthode de Sanger uorophore pour marquer ramorceur (ou primer) dans le capre de la metriude de Dainger. Dans ce cas, les quatre sénes de réactions de séquençage conduisent, comme d'ecrit précédemment, à uano de uas, les quatre sentes de resultoris de sequentivage puritualem, comme desum prevenentim, a quatre families de chaînes d'ADN déposées sur quatre poles parallèles d'un gel d'électrophorèse, qui est quarre tamilles de chaines d'AUN deposées sur quarre potes paraises o un get a électroprovèse, qui est traversé entièrement, suivant sa largeur, par un faisceau laser. Au sein du get, les chaines d'ADN marquées visverse entirement, survent as insigeur, par un tanoceul lisser, au sein au get, es chantes d'autre marquess sont excrées les unes après les autres, dans l'ordre de la séquence, lorsqu'elles passent dans le fasceau

ser, au cours de l'electrophorese. Pour chaque pate d'électrophorèse associée à une famille de réactions donnée, la lumière de fluorescence r-uur chaque pisse u eneccophiorese associée à une lattine de reactions durnée, la authiere de nitre estende est collectée par une optique adaptée et conduite, après filtrage, à un détecteur de type photomultiplicateur laser, au cours de l'électrophorèse, οι contenter par une opinique acaptive si comoune, apres minage, a un oeséctieu de rypé prioremunispicateur. Un avantage offert par cette technique est l'absence de toute partie mobile au sein de l'appareil de detection grâce aux detecteurs fixes qui permettent une détection continue sur les quaire pistes

emultipriorisse Toutefois: l'utilisation d'un seul fluorophore conduit inévitablement à l'utilisation de quatre pistes en ibulerois i utilisation a un seul huorophore condui inevitapiement a i utilisation de quarre pistes en parallèle, sur le gel d'electrophorèse, pour l'étude d'une seule sequence ce qui limité les possibilités d'electrophorese paraleie, sur le get à electrophorese, pour l'etude à une saute seudence ce qui mille les possessines d'évolution sensible et de simplification ultérieures. Il en est de même lors de la détection en temps reel des

naines d'ADN marquees par du priospriore 32. Ces systèmes de détection en temps réel nécessi tent aussi une grande frabilité quant à la comparaison des chaînes d'ADN marquées par du phosphore 32. ues systemes de describin en sema sen necesar (em ausar une y ende casarie visan a le companison ses mobilités electrophorériques d'une piste à l'autre. Or, en pratique plusieurs difficultés peuvent être rencontress en particular, la formation de gradient thermique et la presence d'impuretes au sein du gét pervent induire des distorsions locales des chaînes au sem du gel, et par suite compromettre directement 20

nificultion des cases calls le sequence cretruire. D'autres procédes de détection en temps réel proposent l'utilisation de plusieurs fluorophores afin l'attribution des bases dans la sequence cherchée b sures procedes are unrecarin en remus rem propusent runssands de prosedes monoproves ent d'identifier individuellement chaque famille de chaînes. Ce type de procéde presente l'avantage, dans son u loumnier movimuelement chique lamille de chaines. Le type de procéde présente i avantage, cans son concept d'offrir la possibilité de séparer des chaînes d'ADN issues des réactions de sequençage, sur une concept: a arrivr is possibilité de sépaire des priaires à ANT abside des récentaits de sequences, se une seul et même parte d'électrophorèse. Ces systèmes utilisent une sene de quatre fluorophores, ce qui semble seule et même parte d'électrophorèse. Ces systèmes utilisent une sene de quatre fluorophores, ce qui semble répondre aux divers problèmes soulignés précèdemment. Toutetos, la mise en œuvre spécifique de chacun répondre aux divers problèmes soulignés précèdemment. Toutetos, la mise en œuvre spécifique de chacun

e des systèmes ne renu des appropries que parmenennen reussies. Un premier système a été proposé par L.M. Smith et al. (voir FR.A-2 558 262), dans le cadre des réactions de ces systèmes ne rend ces approches que partiellement réussies un premier systeme a ere propuse par L.m. onnin et al. (von Pri-n-C 200 cod.) usins re-usine uses resolution de sequencage do Sanger. Dans catte technique, les réactions de séquencage doivent s'effectuer dans de sequencage do Sanger. Dans catte technique, les réactions de séquencage doivent s'effectuer dans ue asquencege de beirger. Uais sette technique, les reactions de sequençage doivent s'enectuer dans quatre tubes separés, chacun comprenant un marquieur fluorescent différent les au "primer". L'appareil de quaire usurs separes, criasum comprenent un marqueur nuorescen, content ne au, printer. L'apparent ve détection decrit par Graffin et al utilise une série de fittres interferențiels à bande étroite afin de sélectionner les

longueurs d'onde spécifiques d'émission de chacun des fluorophores. ingueurs à unité succinques à emissaion de crieduri des inuniquirores. À partir de ce procéde relativement simple, le type de système qui en résulte comporte au demeurant Design de de producte consideration ampire, le type de système du ch resulte comporte au derrechent certaines imperfections. En particulier, de séneux problèmes sont engendrés par l'utilisation de molècules

Cette situation se traduit par l'existence de perfurbations au cours de l'électrophorèse dues aux différences fluorescentes de charges et de tailles très différentes. de mobilités électrophorétiques des diverses chaînes marquées differemment. En effet, malgré l'élaboration un informes a equippos anues, so a verses unitares interiores unitares mente. En enel, hauge e escotation conjonite d'un logicel lourd et assez complexe de trafement et de correction des données, ces phénomènes conjonite d'un logicel lourd et assez complexe de trafement et de correction des données, ces phénomènes pervent entraîner certains recouvrements voire des inversions de chaînes au cours de l'électrophorées. empéchant ou même faussant l'identification de certaines bases le long de la sequence

injecutant du meme racosam i nuemonation de certantes baies le long de le saquerce. Par allieurs cette technique requiert obligatoirement l'utilisation de quatre "primers" specifiques marquès Par alleurs cette technique request outigatorement i univación de cuarre, primiers, suscenques inel·dues Ceci entrane, dans les cas ou d'autres vecteurs sont employés, la nécessité d'entreprendre des processas

longs et complexes de synthèse et de purfication de quare nouveaux "primers" fluorescents ngo e i vonumente de symmese et de purmusión de quaire nouveaux "primers muorescents. Un second système proposé par J M. Prober et al. (voir EP-A-Q 252 583) repose, toujours dans le cadre des un secono systeme prupose paru im proper et au type e provulous cesos repuses, toujours une se usere use reactions Sanger, auf futilisation de quatre fluorophores sea aux quarre dOXTP. Ce marquage deterent des reactions banger, sin unissuon de quarre illuorophores les aux quaire daxiv. Le matiques olivient und dXTP procure un certain nombre d'avantages. En particuler, les quatre réactions de séquencage pauvent de la company de la settectuer dans un même tube, un amorceur unque non marque paut alors être utilise et les produits de

réaction peuvent être séparés au sein d'une même colonne d'électrophorèse Septembrie en de septembre au sen is une memo culorine si encrispissere. Cependant, le fait de greffer au doXTP des marqueurs extrinséques, c'est-à-dre les fluorophores, qui sont des molécules asset complexes et tres voluminauses, peu entraîner des décatages imprevisées que soit des molécules asset complexes et tres voluminauses, peu entraîner des décatages imprevisées quals les mobilités de certaines chaînes au cours de l'électrophorèse et surfout restreindre fortement le choix decrymes de ceruleiras. Litantes eu ciudis de renderitariores et sutretir textrendre turentes de decrymes de complémentairé , ceci élimine la possibilité d'utilisation des enzymes courantes comme

amorceur, comme par exemple le fragment de Klenow d'ADN polymerase. Les fluorophores utilises sont

entralement les uenves us le leurrésuent. L'invention à donc pour objet un procédé d'identification des sequences d'un ADN permettant de remèder Linvention a condition or procede didentification des sequences d'un AUN permettant de remoder à la plupart des inconvenients donnés précédemment. Des problèmes similaires apparaissant dans le generalement les dérives de la fluorescéine. a la puppar, una inconvenienta puntes prepatementi, qua propientes aminares apparassant uans le sequençage de tout autre acide nucleique et, en particulier. Cans les ARN, l'invention s'applique aussi a toute

ure mulecule complexe (proteine ou péptide). L'envention à pour objet un procédé d'identification de la séquence des bases d'un acide nucleique. autre molecule complexe (protéine ou peptide).

caractérisé, en ce qu'il comprend essentiellement les étapes suivantes A) production de tragments de l'acide nucleique.
 B) réactions chimiques spécifiques des bases pour former des familles de chaînes de taille différence.

naque retresse comportant a une expremite une usase specifique. Ol marquage spécifique desdites familles de chânes et marquage de chaque nucléotide de chaque chaque famille comportant à une extrémité une base spécifique.

naine a i arce o isotopes ou de perines molecules stables. Di introduction une à une des chaînes marquées dans des moyens d'atomisation et d'ionisation. chaîne à l'aide d'isotopes ou de petites molécules stables.

E) atomisation et ionisation de chaque chaîne marquée à l'aide desdits moyens,

F) selection et détection en temps réel des ions formés par spectromètre de masse, et

r.) selection et detection en temps reel des ions lottines par spactionieure de nasset, et G) traitement des signaux issus de la spectromètrie de masse, pour déterminer la longueur de chaque O transment des signeux issus de la spectrumente de masset, pour determiner la longueur de chaque chaîne en comptant le nombre de nucléotides marques de chaque son détecte et pour identifier le 15

26

60

65

manquese specialque de chaque names de charres Le fonctionnament en temps real est hé au fait que, dans l'invention, les ions sont détectés au tur et à esure de leur tormation. L'onisation des chaînes marquées puis la détection de ces chaînes par un spectromètre ou analyseur de

L'onisairon des chairles marquees puis la detection de cles chaînes per un specifique et d'anierpard de masse facilité considérablement la détection des chaînes marquées et donc l'identification des bases de masse facilité considérablement la détection des chaînes marquées et donc l'identification des bases de l'acide nucléique

Le procéde, selon l'invention s'applique, parfaitement bien au séquençage de l'ADN Le procede seron (invention » apprique parlatement plen au sequencage de l'Auri Ce procede de sequencage fonctionne en continu et ne comporte pas d'étape de radiographie ni Le procede de sequençage ronctiunne en contitu et ne componte pas u enere de readigneme ni d'électrophorese selon l'art anténeur. En outre, les chânes de nucléotides marquées sont examiness et

registres une par une et nun guiussement secon i et ameniu. Selon l'invention, on effectue un marquage de tous les nucléotides de chaque chaîne, la détection des analysées une par une et non globalement seion l'art antérieur Seton i invention, on enectue un marquage de tous les nucleonides de chaque chaine : la désection des marqueurs de chaque nucléotide donne l'information de la longueur de ces chaînes alors que la détection du marqueurs de chaque nucléotide donne l'information de la longueur de ces chaînes alors que la détection du

arquage specifique conne i inscrimation de la tamilie de chaînés. Enfin, l'ionisation et la spectromètrie de masse permettent l'utilisation de marqueurs liés de manière marquage spécifique donne l'information de la famille de chaînes entin, Homsalion et la spectrometrie de masse permettent lutilisation de marqueurs lies de manière intrinseque du extrinseque aux différents nuclèotides, ce qui autorise l'application de ce procédé aux diverses intrinseque du extrinseque aux différents nuclèotides, ce qui autorise l'application de ce procédé aux diverses ninchaeque ou extinuación des unicientes hobroriumes. Le qui autoria e autoria el expendient de ce provede aux unicientes methodes chimiques de séquencage les plus répandues, et en particuler, celle de Sanger et de ses variantes.

winder uninques de sequençaiga les plus repardues, et et parturair, delle de sanger et de ses variantes. De laçon avantageuse, le marquage spécifique de chaque famille de chaînes et lou le marquage de tous les Le raçon avantagense, le manufuege specinique de disdue anime de chantes envolve intellueged de nucleotides de chaque chaîne sont effectues avec des isotopes ou des petites molècules stables.

pueditives de chaque chame sont enecutes avec des politiques du des pentes transculers avaites. Par petites molècules, il faut comprendre des composés chimiques pouvant être liès aux nucleotides et qui r air presses insurculates, in rant comprehendre des composes criminques pouvant erre lies aux nucleotides et qui comportent au moins un element marqueur identifiable par spectromètrie de masse, tel que CF<sub>2</sub> S(CH<sub>3</sub>)s, en Comporterni au moins un element mai queur loemmeure par spécimente de masse se que cm², pilichtaja, en Unitisant par exemple, une technique decrite dans l'arricle Tetrahedron, vol. 25, № 40, pp. 4915-4918 (1985).

Le procede selon l'invention, grâce à l'utilisa tion de trois, quatre ou pluseurs isotopes ou petites molecules. Les Nouveus aeurs avenues, grace a l'umas averse avenue, quane un projette à soutpets du petites moietures atables c'est-à-dire non radioactives, donc non dangereuses pour les operateurs, facilité considérablement C.D. Pein, D. Cech.

s interventions numeries care le procéde didensification. L'invention à aussi pour objet une installation pour l'identification des bases d'un acide nuclèique les interventions humaines dans le procéde d'identification.

une soutre de latimes de claimes de plases de plases

moyen d'isotopes ou de petites moiecules stables. - ues moyens o anomesaron de unaque chame melqueer. - des moyens pour introdure une à une dans les moyens d'atomisation les chaînes marquées,

עם אוויקאיים עם אוויקאיים עם אוויקאיים אוויקא

 un spectrometre de masse equipe de plusieurs detrectieurs pour selectrometre et uevectet est sons tormes et des moyens de tratement des signaux fournis par les défecteurs pour déterminer la longueur de chaque - ues integreta de instrumento des arginatas notarios par les describas pour detendiren la longueur de chaque chaîne en complant le nombre de nucléotides marqués de chaque ion détecté et identifier le marquage

perinque de chaque tamille de chaines. Comme on l'a dit précédemment, l'invention s'applique à toute autre molècule complexe que l'AON et TARN, Aussi, l'invention a encore pour objet un procédé d'identification des motifs d'une molécule organique Complexe ou d'un fragment de cette molècule, comportant essentiellement les étapes suivantes

marquage speurique uesquis mutils.
Introduction un à un des motifs marquès dans des moyens d'atomisation et d'ionisation.

- atomisation et ionisation de chaque motif marque à l'aide desdits moyens.

detection en temps réel des ions formes par spectrometrie de masse, et

- uertestion en remps rem oes soms normes par spectrometrie de masse; et - tratement des signaux issus de la spectromètrie de masse pour déterminer la longueur de la protèine en comptant le nombre de motés marquès et pour identifier leur marquage spécifique umpains se numure de mutes marques es pour identimes seul manquage specimique. D'autres caracteristiques et avantages de l'invention ressortiront de la description qui suit, donnée a titre

illustratif et non limitatif aux dessins annexés, dans lesquels : - la figure 1 représente un schema d'ensemble des étapes de séquençage automatisé d'ADN selon

l'invention - la figure 2 représente un schéma d'ensemble d'une installation automatisée pour le séquençage

d'ADN conforme à l'invention. - la figure 3 représente le type de données produites par un séquençage d'ADN selon l'invention La description qui va suivre se réfère au sequençage de l'ADN, bien que comme on l'a dit précèdemment.

l'invention s'applique à tout acide nucléique (ARN par exemple) et même à toute molécule organique complexe (proteine, peptide).

La première étape du procédé, en référence à la figure 1, consiste à fragmenter un génome d'ADN (étape a). conformément à l'art antérieur cité. Cette étape de fragmentation porte la référence 2 sur la figure 1.

Les chaînes obtenues sont alors soumises aux réactions de séquençage (étape b) selon les méthodes de Sanger, si ce n'est que l'échelle peut être augmentée en cas de besoin pour fournir une intensité de signal adéquate pour la détection de chaque famille de chaînes. Aucun nucléotide radio marqué n'a besoin d'être inclu dans les réactions de séquençage. Les réactions de sequençage sont symbolisées par le bloc 4 sur la

Selon l'invention, tous les nucléotides incorporés selon Sanger au cours de la synthèse des chaînes figure 1 nucléotidiques complémentaires du fragment d'ADN à sequencer sont marquès ou bien comportent une caractéristique identifiable par spectromèrie de masse. En outre, un marquage spécifique de chaque famille

de chaînes issue des reactions de séquençage doit être effectué Le marquage des chaînes nucléotidiques, symbolisé par le bloc 6, peut être réalisé en utilisant un isotope ou une combinaison d'isotopes stables, constituants éventuels de petites molécules stables, afin de marque specifiquement chaque nucleotide de chaque chaîne ainsi que l'extrémité de ces chaînes en fonction de la

base terminale de ces chaînes Le marquage spécifique de la famille de chaîne peut être réalisé sur l'amorceur, comme décrit dans le document FR-A-2 558 262, ou sur le didesoxy nucléotide, comme décrit dans le document EP-A-0 252 583. mais en utilisant des isotopes ou petites molécules stables

Grâce à ce double système de marquage, chaque chaîne d'ADN issue des réactions chimiques de séquençage comporte une double information | 1) la détection du nombre de marqueurs et donc du nombre de nucléotides constituant chaque chaîne, qui donne l'information taille de la chaîne (comme le faisait anteneurement l'électrophorése -étape d-), 2) la détection du marqueur stable spécifique de la famille qui donne le type de nucléotide, et donc de base, existant à l'une des extrémités de la chaîne (comme le faisait antérieurement les fluorophores ou les isotopes radioactifs -étape c-).

Le marquage isotopique selon l'invention des chaînes d'ADN peut être realisé au niveau de l'un des trois composes chimiques constituant les nucléotides, c'est-à-dire la base, le sucre ou le groupement (riphosphate ou éventuellement sur un groupement alpha-thiophosphate dans lequel l'atome d'oxygène lie au premier

phosphore est remplace par un atome de soufre Le marquage des chaînes peut être réalisé de façon intrinsèque sur le carbone, l'hydrogène, l'azote l'oxygene ou même le soutre entrant dans la constitution des trois composés chimiques d'un nucléotide L'utilisation d'isologes naturels stables selon l'invention, offe un chox, plus important que les isotopes radioactifs genéralement utilisés (22P, 35g, 14C, 3H) pouvant entrer dans la constitution des différents

composes nucleotidiques Une liste de ces isotopes stables, incluant leur abondance naturelle relative, est donnée dans le tableau 1,

Tableau 1 isotopes naturels stables pour marquage intrinseque des composés nucléotiques

45

	Isotopes	Abondance relative (
50	D	0,015
	130	1,10
	15N	0,37
	170	0.038
55	180	0.20
	325	95.02
	33S	0,75
	345	4,21
	365	0.02
60		

Du point de vue chimique, les réactions pour incorporer un isotope stable ou un isotope radioactif sont les mémes

Les divers procédés de marquage de molecules biologiques par des isotopes stables existent et sont 65 largement repandus (voir à cet effet l'article "The preparation of purines, pyrimidines, nucléotides and related

compounds labelled with "O and other stable isotopes", E. Petreanu, J.S. Cohen and D. Samuel, Proceedings compounds lapeled with "10 and other stable systopes". C. Ferreanu, 3 o Johns and Compounds, pp. 465-497) of the 2nd International Conference on methods of preparing and storing labeled compounds, pp. 465-497). of the 2nd International Conference on methods of preparing and storing labeled compounds, pp. 489-497).
Differentes molécules marquees, en particulier les nucléotides peuvent être produites par voie de synthèse ou Universities molecules marquees, en particulen les nucleotides peuven etre produites par vue de syminése out par voie enzymatique. En particuler, de nombreux composés nucleotidiques marques avec des isotopes pai, voie enzymatique, en particulier, de nombreux composes nucleologiques marques avec das isotopes stables sont disponibles sur le marché pour des applications en biologie et en pharmacologie differentes de stables sont disponitions sur le marche pour des applications en viologie et en priermativolger unterentes de celes de l'invention (voir par exemple l'article Dehalogeration reactions in fast atom bombardment

was-spectremetry, Amel Onem, 1909-90, 1910-771). Des exemples de nucléotides marqués à l'aide d'isotopes naturels sont donnes dans les annexes I et il mass-spectrometry, Anal. Chem, 1984.56, 1975.77). 

nvention. Comme données dans ces annexes I et II, le marquage peut être effectué sur les formes Comme connees cans ces annexes i et ii, ie marquage peut etre emectue sur les formes désoxytrohosphate (annexe i) ainsi que sur leurs analogues alpha-thiotrohosphate (annexe ii) pour le

desoxytronosphate (annexe I) ainsi que sur leurs analogues alpha-miotrophosphate (annexe II) pour le marquage de chaque nucléotide et peut être effectué sur les formes didésoxytriphosphate (annexe I) ainsi que marquage de chaque nucleone et peut etre emectue sur les normes dibesoxytriphosphate (annexe i)? sur leurs analogues alpha-thiotriphosphate (annexe ii) pour le marquage spécifique de la famille ul reurs amalogues alpha-iniouriphosphate (armexe III) pour le marquage specimique de la tamine. Comme autre nucléotide marque intrinséquement utilisable pour l'information de taille des chaînes issues

Comme autre nucleotide marque intrinsequement utilisable pour finitiomation de taille qes chaînes issues des réactions de Sanger, on peut citer le 1 ou 3-15 NdATP, l'q-16 OdTP, le 4-16 OdTP, le 2-13 CdATP ou dCTP. des reactions de Sanger, on peut citer (e. 1 ou 3-14NoA I.P.; to 1400CI.P.; to 4-900 I.P.; to 2-900A.P. ou observing le 8-12CAATP, ito 5-12CAATP, ito 5-12CA

merny: "UOI!".
Il est aussi possible, selon l'invention, d'effectuer un marquage extrinsèque des nucléoildes de chaque Il est aussi possible, selon i invention, d'effectuer un marquage extrinseque des nucuenties de chaque chaîne, des didésoxy nuclèotides de ces chaînes ou même des amorceurs. Ce type de marquage comporte le methyl 13CdTTP. chaine, des didesoxy nucleologes de ces chaines ou melme des amoriceurs. Le type de marquage comporte l'avantage de pouvoir lier au nucléotide des atômes (sous forme mono ou polyatomique) qui ne sont pas ravantage de pouvoir ser au nucleotide des atomes (sous torme mono ou polystomique) qui ne sont pas naturellement presents dans ces molécules. Ces marqueurs sont constitués d'élements ou de petits

complexes moleculaires d'eléments sicalins, alcalino-terreux ou halogéniques unpreses indiculaires o vientetts dicipino, piccellu-terreux ou narogeniques. Dans le cadre des réactions de séquençage, le marquage extrinsèque selon l'invention présente moins de uans le cadre des reactions de sequençage, le marquage extrinseque selon invention presente moins de risque de dénaturer les proprietes physiques et chimiques des nuclèotides, contrairement à l'adjonction selon insque de denaure les propriètes prigraudes et chimiques des nucleondes, contrarement à l'adjonction selon. Fait anténeur de grosses molécules telles que les chromophores ou fluorophores qui ne peuvent être lièes. Tari amerieur de grusses mojecules tejes que les um unupritures du hubrippiores qui ne peuvent etre iless que par l'intermédiaire de bras spécifiques relativement complexes. Ains), les éléments ou petits complexes que par intermediaire de cras specifiques relativement complexes. Ainsi, les élements ou pents complexes molèculaires selon l'invention peuvent être directement greffés au nucleotide en profitant des liaisons. moleculaires selon i invention peuvent etre directement grettes au nucleonale et prumant use nessona covalentes disponibles au sem de la base ou du sucre de chaque nucléotide (voir par exemple l'afficile d'Anal

niern. Lite en gernier). Des exemples de marquage extrinsèques selon l'invention sont donnés en annexe ill. Cet exemple donne ues exemples de marquage extrinséques selon tinvention sont donnés en annéxe III. Let exemple donné un jeu complet de bases nucléofidiqués de l'ARN et ADN marquées extrinséquement. En particulier, X Chem cité en dernier).

30

24

55

presente un atune de tiudi, de culore, de unone du o roce. Comme nucléotide marque par des halogènes utilisables dans l'invention, on peut citer le 8-bromo dGTP, le représente un atome de fluor, de chiore, de brome ou d'iode. S-bromo dCTP, le 3-bromo chioro ou lodo dTP, le 5-bromo iodo ou fluoro dTP le 3-FdGTP cromo dulh, le 3-promo cnioro ou logo d'1 P, le p-bromo lodo ou fluoro d'1 P. le 3-positif En tenant compte de l'existence de plusieurs isotopes stables possibles des differents marqueurs en ienani cumpite de l'existence de plusieurs isotopes stables possibiles des différents marqueurs extrinséques (élèments haloganes par exemple), les possibilités de marquage isotopiques des nucleotides

ont rorrement acciues. Une liste des différents isotopes stables d'éléments halogénes est présentée dans le tableau 2 sont fortement accrues accompagnès de leur abondance naturelle relative.

# Tableau 2

24 23

35CI

	Tableau 2	4
isotopes naturels :	stables des éléments halogènes narquage extrinsèque	
	Abondance relative (%)	
isotopes	100	,
19F	75.77	

a7Cl	50,69
79Br	49.31
€¹Br	100
127	marquages multiples des nucléotides de chaque
Conformement à	l'invention, il est possible d'effectuer des marquages multiples des nucleotides de chaque réactions de séquençage (information taille), des didesony nucleotides et amorceurs réactions de séquençage (information taille), des didesony nucleotides et de montages to de ce chaînes ces marquages étant des marquages intrinséques et/ou extrinséques la de ces chaînes ces marquages étant des marquages intrinséques le 13-aille AGTP, le

chamie issue des reactions de sequençage (miormation iame), des undestay indicatores et col extinsequi (information (smille) de ces chaînes, ces marquages étant des marquages intrinsèques et cu extinsequ normation ramite), de ces chaines, ces marquages etant des marquages intrinseques e/ou eximiseques. En premier lieu, on peut citer des possibilités d'incorporer plusieurs isotopes marqueurs de même type sur en premier leu, co peut cher des possumeres a incorporer pluseurs biocopes marqueurs de neine nye sur un nucleoted donné. Comme nucléotide marqué de ce type, on peut citer par exemple le 1,3-18N dATP, le un nucièciate conne. Comme nucièciate marque de ce type, on peut citer par exemple le 1,3-18N DAIP, le 1,3-18NGCP, le 2,8-dichioro ou dibromo ou diffuoro dATP, le 5,5-dibromo dCTP. Ceci permet d'accroître

1.3- "NOUTP, le 28-dichioro ou apromo ou amuro unit, le 3.5-aprilire de l'identification des chaînes nucleotidiques d'autant la sensibilité obtenue lors de la détection et de l'identification des chaînes nucleotidiques autant la sensionité obtenue lors de la détection et de l'identification des chames hubeutituiques. En second lieu plusieurs marqueurs différents peuvent être incorporés au sein d'un même nucléotide afin ch second neu, prosieurs marqueurs omerents peuvent etre micropores au sein o un meme nucleorore ann d'affiner sa signature. De plus, la detection simultanée de plusieurs marqueurs augmente considérablement les combinaisons de marquages possibles

L'exemple suivant montre que trois types de marqueurs isotopiques stables suffisent à l'identification de quatre nucleotides différents :

totide o	a type	Α
éotide d éotide d	te type	G
	éotide (	éctide de type léctide de type léctide de type

.

55

60

Ainsi, les marquages intrinsèques et extrinsèques des nuclèotides à l'aide d'isotopes stables peuvent être

combinés et offrent de réelles possibilités multiples et variées. Cette methode de marquage s'avère plus souple et plus nche d'évolution que les marquages radio-isotopiques ou par fluorophores selon l'art antérieur qui restent forcément plus limités dans leurs possibilités. De plus, la plupart des inconvénients rencontrés avec ces derniers lors de leur manipulation et de

leur utilisation sont évités voire complétement éliminés. Les chaînes d'ADN marquées sont alors placées dans une source d'échantillon 8 appropriée à leur introduction individuelle dans une source d'ions. Cette étape est réalisée grâce à la formation de

microgouttelettes contenant au plus une molécule marquée. Celles-ci sont ensuite introduites successivement une à une dans une source d'ions chargée de

20 l'atomisation et de l'ionisation. Cette étape porte la référence 10 sur la figure 1 et correspond à l'étape D et E Les ions formés pour chaque chaîne marquée sont alors sélectionnés puis détectés par spectromètrie de masse, comme indiqué en 12, et les signaux issus de la spectrométrie de masse sont traités, comme indiqué en 14 en vue de reconstituer l'ADN de départ (étape G).

L'installation permettant le séquençage automatisé de l'ADN, conformément à l'invention, est schema tisée sur la figure 2 La fragmentation du genome, les réactions chimiques de séquençage et le marquage étant effectues de laçon discrète ou de façon indépendante dans le temps, ne sont pas représentés sur la figure 7 L'installation représentée comporte une source 18 de chaînes marquées d'ADN a l'aide d'isotopes stables seion l'invention, placée sous pression et contenant une cuve 20 destinée à recevoir l'échantillon 22 des

chaines marquees d'ADN diluees à environ 10-19 mole/l dans une solution tampon. Un tube capillaire 24 d'environ une dizaine de micron de diametre plongeant dans l'échantillon 22 assure l'introduction une à une des chaînes d'ADN marquées dans une source d'ions 26. Le cristal pièzoélectrique 28 permet le fractionnement homogene la nébulisation en microgouttelettes de quelques micrometres de

Elles sont alors introduites une à une dans la source à plasma 32 équipée de bobines électromagnétiques diametre de l'affluent du tube 24. 33 reliess à un genérateur radiofrequence 34. Dans la zone 32a la plus chaude du plasma, où la temperature

atteint 7000 à 8000°K, les molècules d'ADN sont atomisées et ionisées. La torche à plasma est la source préfèree d'ionisation selon l'invention car elle permet l'atomisation complète des molécules et induit une simplicité maximale dans le spectre de masse obtenu après analyse

Les ions formes sont alors pris en charge par l'optique d'entrée 36 d'un spectromètre de masse 36 qui peut d'elements simples être du type à secteur magnétique où un vide différentiel s'installe grâce à des pompes primaires et secondaries 39 ; la pression au niveau de l'optique io nique 36 est de l'ordre de 10°4 torr (10°2Pa) et la pression

au niveau des pièces polaires 40 du spectromètre est de l'ordre de 10<sup>-5</sup> torr (10<sup>-4</sup> Pa). Le premier rôle du spectromètre de masse 38 est de sélectionner et d'identifier grâce à son optique ionique 36 et ses pièces polaires 40, la présence des masses spécifiques obtenues grâce au traceur isotopique utilise précèdemment. Le spectromètre de masse 38 peut opèrer en mode sélectif ce qui consiste à balayer uniquement les masses caractéristiques dues à la présence du marqueur isotopique, afin de fournir un

La détection par spectromètrie de masse permet à la fois de détecter les chaînes d'ADN marquées au fur et maximum de sensibilité à mesure de leur sortie de la source de tragments 18, d'identifier, grâce à leurs marqueurs spécifiques, à quelle famille elles correspondent ainsi que de déterminer leur taille grâce au comptage du nombre de

Les moiecules ionisées et sélectionnées par le spectromètre de masse sont alors détectées à l'aide d'une nucleotides serie de détecteurs 42 situés à la sortie du spectromètre 38, du type multiplicateur d'électrons et qui sont sensibles a l'impact d'un seul ion. Les ions sélectionnés peuvent donc être détectés individuellement, ce qui favorise l'obtention d'une sensibilité maximale de l'installation totale de séquençage.

Le signal de comptage fourni par chaque détecteur 42 est alors reçu dans un micro-ordinateur 46 du type PC qui gère automatiquement, à la fois le pilotage des divers sous-ensembles (18, 26, 38) de l'installation, le

traitement et la représentation des données. Pour chaque chaîne marquée, l'événement correspondant à son identification par spectromètrie de masse est caractèrise par la détection simultanée : 1) d'un marqueur caractérisant la famille de la chaîne, 2) d'un certain nombre d'autres marqueurs associes à chaque nucléotide de la chaîne. Ce nombre est directement relié à la quantite de nucléotides constituant la chaîne, donc à sa longueur (autant de marqueurs que de

Au bout d'un certain temps, de nombreuses chaînes sont passées une à une dans l'ensemble de détection. nucléotides par exemple)

Une statistique est alors constituée afin d'établir des diagrammes tels que représentés sur la figure 3. Les une statistique est alors constituee afin d'établir des diagrammes tels que représen diagrammes de la figure 3 correspondent à l'ADN hypothètique TGACTGCACGTA diagrammes de la figure 3 correspondent à l'ADN hypothètique

ingrammes de la ngure y correspondent a l'AUN hypometique l'UAUIGUAUGIA. Le nombre d'évenements associés à la présence d'un marqueur specifique stable d'une famille (sur le Le nombre o evenements associes a la presence o un marqueur specimque basine u une familie (sur le didésoxy par exemple) est porté en ordonnée : isotops 1 pour les chaînes terminées par l'adenne (tamile A) didesoxy par exemple) est porte en ordonnee : isolope : pour es chaines terminees par l'apenne (tamilie A) isolope 2 pour les chaînes terminées par la guarine (tamilie G) : isolope 3 pour les chaînes terminées par la guarine (tamilie G) : isolope 3 pour les chaînes terminées par la isologe 4 pour les criaines terminées par la guannie (tamue u), isologe 3 pour es chaines terminées par la cytosne (famille C) et isologe 4 pour les chaînes terminées par la thymine (famille T). L'unité des ordonnées

sc arbitraire. Le nombre de marqueurs stables associés aux nucléotides et détectés simultanément lors des événements Le nombre de marqueur s'apoles associées aux incidentes et dérectes serunianement lors des evénements est porté en abscisse. Ce nombre est directement relle (voire proportionnes) au nombre de nucléotides (n est porte en appoisse. Le nombre est directement rete (voire proportionner) au nombre de nucleotides (n. n.-1, n.+2... avec n variant de 1 à 400 anviron) constituant les chaînes détectées dans l'ensemble du système

e detection. Les diagrammes établis pour les quatre types de familles donnent alors les informations nécessaires pour de détection.

erennmer rensemble de la sequence de l'Aura de depen. À titre illustratif, on donne, ci-après, trois exemples de marquage des différentes chaînes de l'ADN déterminer l'ensemble de la séquence de l'ADN de départ 15

pointendue Tual Tual Au III.
Au cours des réactions de séquencage selon Sanger, tous les nucléotides libres présents (désoxy et Au cours des reactions de sequençage selon banger, tous ses nucleotides notes presents (desoxy et didésoxy) sont marquès. Leur incorporation dans les bras synthètisés engendrent des chaînes de la forme hypothétique TGACTGCAACGTA

désoxy nucléotides marqués (X)

didesoxy nucleotides EXEMPLE 1 desoxy nucleotides marques specifiquement

marques marques au 28 dichloro ddATP (35Ci) 13c 5.5 dibromo ddCTP (79s.) 2 - 13C dATP

8 - 13C dGTP 8.3' diffuoro ddGTP (19r) 5.3' diiodo ddTTP (127)) 2 - 13C dCTP methyl 13C dTTP

35

40

45

40

5	exemp désoxy marque α 32S α 32S α 32S α 32S α 32S α	nuclée is au 3 dATP dGTP dCTP	otides 2S	1,0 2,0 2,0	m spéci 15N c 13C c 13C c	dCTP							
10													
	Exem	ples	₫e	chai	nes	marq	uées	. :		_			-
	т	G	A	C	т	G_			Т	G	A .	c	'\
15	1	i	1	1	1	17	0		1	1	1		100
	32 <sub>S</sub>	32 <sub>S</sub>	32 <sub>S</sub>	32 <sub>S</sub>	32 <sub>S</sub>				32 S	32 <sub>S</sub>	32 <sub>S</sub>	32 ş	
20	т	G	A	С	τ	G	С	Α_					
	ĺ	1	1	1	1	1	1		15 <sub>N</sub>				
	32 <sub>5</sub>	32 <sub>S</sub>	32 <sub>S</sub>	32 <sub>S</sub>	32 <sub>S</sub>	32 <sub>S</sub>	32 <sub>S</sub>						
25	т	G	А	С	т	G	с 🛴						
	i	1	ï	1	1	1	_	13 с					
30			32 <sub>\$</sub>			32 <sub>S</sub>							
35	déso	MPLE 3 xy nucl ues au	éctides	5		xy nuc marqué cifiquen	s	s					
40	8 bro	amo de Ob omo Ob omo Tb omo	TP TP		α- 33S α- 34S	ddATP ddGTP ddCTP ddTTP							
45													
50													



L'invention ayant été décrité en détail, à est évident que l'ARN ou d'autres types de molécules organiques L'invention ayant ene deutrie en desait it est evident que l'Ann ou d'autres types de molècules organ-complexes organiques pauvent être analysées grâce au procédé et à l'installation selon tinvention.

Par exemple en marquant chacun des acides aminés d'une proteine et plus spécifiquement certains d'entre Har exemple en marquant chacun des acides amines d'une proteine et plus spécifiquement certains d'entre eux. « est possible, grâce à l'invention de compter le nombre d'atomes marqueurs et d'en dedure le nombre eux. n'est possible, grace à invention de comprer le nombre à atomes marqueurs et d'en dedure le nombre d'acides aminés de la proteine ainsi que le nombre des acides aminés qui avaient été marques plactures amines on is prutiente ams que le nombre des sudes animes qui avaient ete marques specifiquement. En renouvelant cette operation pour tous les types d'acides amines on peut déterminer la

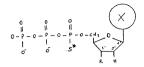
omposition en acioes arrines de la proteine. De même lacon, il est possole de déterminer la longueur des fragments de restriction d'ADN et d'ARN en composition en acides aminès de la protèine. ue meme tacon, il est pussable de déterminer la longueur des tragments de restriction d'AUN et d'AR-Comptant le nombre de marqueurs, celui-ci étant égal au nombre de nucleotides desdits fragments.

35

6. 18 (d) dTTP

ANNEXE I 

## ANNEXE II



55

## ANNEXE III

5 10

15

20

25

30

45

50

55

60

65

URACIL

GUANINE

CYTOSINE

## Revendications

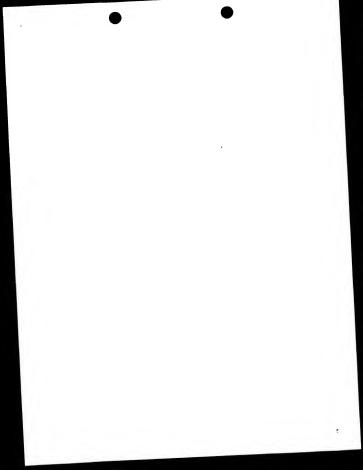
- Procèdé d'identification de la séquence des bases d'un acide nuclèique, caractèrisé en ce qu'il
- comprend essentiellement les étapes suivantes : A) production (2) de fragments de l'acide nucléique. B) réactions chimiques spécifiques (4) des bases pour former des familles de chaînes de taille
  - différente, chaque famille comportant à une extremité une base spécifique, C) marquage spécifique (6) desdites familles de chaînes et marquage de chaque nucléotide de
  - chaque chaîne à l'aide d'isotopes ou de petites molécules stables, D) introduction (24) une à une des chaînes marquées dans des moyens d'atomisation et
    - d'ionisation (26, 28, 30, 32). E) atomisation et ionisation (10) de chaque chaîne marquée à l'aide desdits moyens.
    - F) sélection et détection (12) en temps réel des ions formés par spectrometrie de masse, et G) traitement (14) des signaux issus de la spectromètrie de masse, pour determiner la longueur de
  - chaque chaîne en comptant le nombre de nucleotides marqués de chaque ion détecté et pour

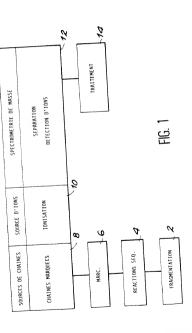
### ED 0.360 677 A1

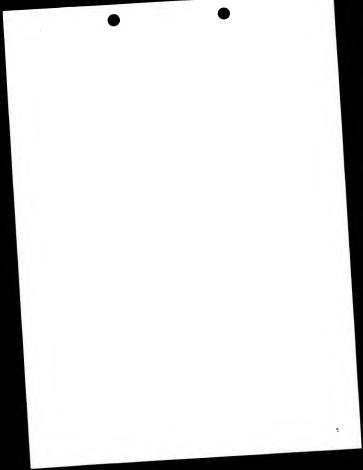
- identifier le marquage spécifique de chaque familie de chaînes. 2 Procédé selon la revendication 1, caractérise en ce que le marquage est réalise de façon intrinseque avec au moins un isotope choisi parmi D, 13C, 15N, 17O, 15O, 32S, 33S, 34S et 35S
- avec au moins un isotope chois parm, D. <sup>14</sup>C. <sup>18</sup>N. <sup>14</sup>C. <sup>14</sup>C. <sup>18</sup>S. <sup>14</sup>S. <sup>14</sup>S. <sup>14</sup>S. <sup>18</sup>S. <sup>18</sup> Procècé selon l'une quelconque des revendications 1 à 3, caractèrise en ce que l'acide nucleique est
  - 5 Procèdé selon l'une quelconque des revendications 1 à 4, caractérisé en ce que l'étape B) consiste à

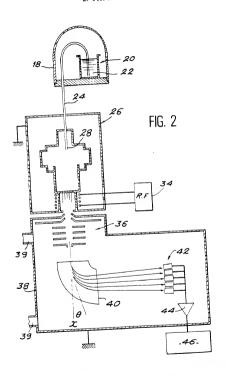
5

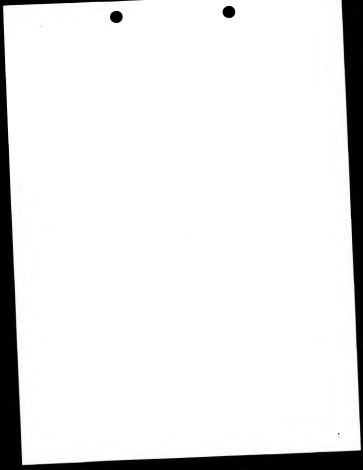
- synthètiser une chaîne de nucléotides complémentaire d'un fragment dudit acide nucléique à partir d'un amorceur jusqu'à ce qu'elle incorpore un didésoxynucléotide. 6. Procédé selon la revendication 5, caractérisé en ce que le marquage est effectué sur un élément
- choisi parmi la base, le sucre et le triphosphate du didésoxy nucléotide et l'amorceur
- Procédé selon la revendication 5 ou 6, caracterisé en ce que chaque nucléotide incorporé au cours de la synthèse de chaque chaîne est marqué 8. Installation pour l'identification des bases d'un acide nuclèique, comportant essentiellement
- une source (18) de familles de chaînes de bases de longueur differente, chaque famille de bases étant spécifiquement marquée, et les nucléotides de cheque chaîne étant marqués, ces marquages étant taits
- des moyens (22, 24) pour introduire une à une dans des moyens d'atomisation les chaînes marquées. au moyen d'isotopes ou de petites molècules stables.
- des moyens (26, 28, 30) d'atomisation de chaque chaîne marquèe.
- des moyens (32) d'ionisation des chaînes individualisées et atomisées. - des inoyens (ac) undividuel des claimes maintaines de la laboration de la defecter les - un spectromètre de masse (38) équipe de plusieurs détecteurs (42) pour selectionner et détecter les
- e. des moyens (46) de traitement des signaux fournis par les détecteurs (42), pour déterminer la longueur des moyens (46) de traitement des signaux fournis par les détecteurs (42), pour déterminer la longueur de chaque chaîne en comptant le nombre de nucléotides marqués de chaque lon détecté et pour de chaque chaîne en comptant le nombre de nucléotides marqués de chaque lon détecté et pour
- identifier le marquage specifique de chaque famille de chaînes. 9 installation selon la revendication 8, caractèrisée en ce que les moyens d'ionisation comportent une
- iorine à piasitia (92). 10. Procéde d'identification des motifs d'une molécule organique complexe ou d'un fragment de cette torche à plasma (32)
- 20 molécule comportant essentiellement les étapes suivantes
- -marquage spécifique (6) desdits motifs par des isotopes ou petites molècules stables Introduction un a un des motifs marqués dans des moyens d'atomisation et d'ionisation (26, 28, 30, 32)
- atomisation et ionisation (10) de chaque motif marque a l'aide desdits moyens
- detection (12) en temps réel des ions formés par spectromètrie de massé, et -traitement (14) des signaux issus de la spectromètrie de masse pour déterminer la longueur de protèine 35 en comptant le nombre de motifs marquès et pour identifier leur marquage spécifique



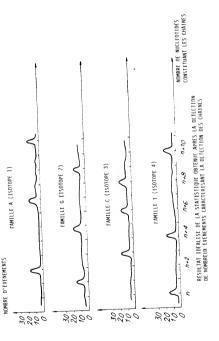


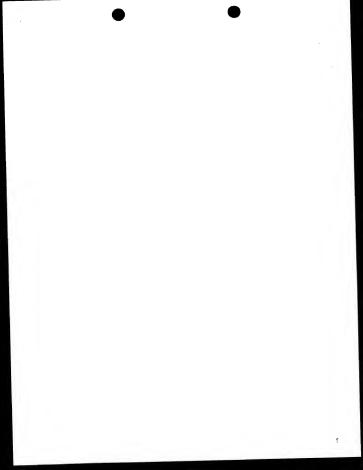






FG. 3







## RAPPORT DE RECHERCHE EUROPEENNE

humero de la demande

EP 89 40 2549

DC	CUMENTS CONSIDE	RES COMME PERTINE	NTS		
Catégorie	Citation du document avec i des parties per		Revendication concernee	CLASSEME DEMANDE	T DE LA
D,Y	FR-A-2 558 262 (CA OF TECHNOLOGY) * En entier *	IFORNIA INSTITUTE	1-13	C 12 Q	1/68
Y	DE-A-3 312 929 (GE BIOTECHNOLOGISCHE F * En entier *	SELLSCHAFT FÜR DRSCHUNG mbH)	1-13		
Y	US-A-4 022 876 (M. * En entier *	ANBAR)	1-13		
Α	WO-A-8 505 642 (K.	RODLAND)			
A	PATENT ABSTRACTS OF 111 (P-451)[2168], JP-A-60 242 368 (HI K.K.) 02-12-1985	JAPAN, vol. 10, no. 25 avril 1986; & TACHI SEISAKUSHO			
				DOMAINES T RECHERCHI	ECHNIQUES S (Int. CL5)
				C 12 Q G 01 N	
Lep	résent rapport a été établi pour to	utes les revendications  Date d'administration la reclarecte		Expublica	

CARTAGENA Y ABELLA P. 01-11-1989 LA HAYE

### CATEGORIE DES DOCUMENTS CITES

- X | particulierement pertinent a lus seul
  Y | particulierement pertinent en combination avec un
  autre éocument de la même categorie
  A | arriere plan technologique
  O | divulgation non-ecrite
  P | document intercaliaire
- Y: theorie ou principe a la base de l'Iavention E: document de brevet anterieur, mais publié a la date de dépôt ou après certe date D: cité dais la démande L: cité pour d'autres raisons

  - 4 : membre de la même famille, document correspondant

